

CLAIMS

What is claimed is:

1. A method of content addressable data storage and compression for computer memory comprising:
 - 5 providing a chunk of data comprising a quantity of input data;
 - retrieving a memory block from computer memory;
 - searching for a segment of the chunk that matches the memory block; and
 - 10 if a matching segment is found:
 - discarding the matching segment and providing a retrieval key for the memory block as a retrieval key for the matching segment;
 - 15 identifying an unmatched portion of the chunk that does not match the memory block; and
 - storing the unmatched portion and providing a retrieval key for the unmatched portion.
 - 20
2. The method of claim 1 wherein providing a chunk comprises retrieving from input a quantity of input data for storage and compression having a chunk size larger than a maximum memory block size.
3. The method of claim 1 wherein retrieving a memory block from computer memory comprises retrieving from computer memory a memory block having a memory block size no greater than a maximum memory block size.

4. The method of claim 1 wherein searching for a segment of the chunk that matches the memory block comprises searching at a repeating memory interval through a search section of the chunk for a segment of the chunk that matches the memory block.
- 5
5. The method of claim 4 wherein the memory interval is one bit.
6. The method of claim 4 wherein searching at a repeating memory interval through a search section of the chunk for a segment of the chunk that matches the memory block comprises:
- 5 calculating a weak checksum for the memory block;
- calculating weak checksums for segments of the search section of the chunk;
- comparing the weak checksums for the segments with the checksum for the
- 10 memory block; and
- if a segment is found with a weak checksum equal to the weak checksum of the memory block:
- 15 calculating a strong checksum for the memory block;
- calculating a strong checksum for the segment with the matching weak checksum;
- 20 comparing the strong checksum of the memory block and the strong checksum for the segment with the equal weak checksum;
- determining that the search has found a segment having contents that match

25 the contents of the memory block if the strong checksum of the memory block
and the strong checksum for the segment with the matching weak checksum
are equal.

7. The method of claim 6 wherein calculating a strong checksum for the memory
block comprises calculating a static strong checksum for the memory block.

8. The method of claim 6 wherein calculating weak checksums for segments of
the search section of the chunk comprises calculating rolling weak checksums.

9. The method of claim 1 wherein storing the unmatched portion of the chunk
comprises storing the unmatched portion of the chunk as a new memory block
having a memory block size equal to the size of the unmatched portion of the
chunk.

5

10. The method of claim 1 wherein providing a retrieval key for the unmatched
portion of a chunk comprises:

calculating a weak checksum for the unmatched portion of the chunk; and

5

calculating a strong checksum for the unmatched portion of the chunk.

11. The method of claim 1 wherein providing a chunk comprises:

retrieving from input a quantity of data equal in size to the sum of the sizes of
the matching segment and the unmatched portion; and

5

concatenating the retrieved input quantity to a remaining portion of the chunk
that remains after discarding the matching segment and storing the unmatched
portion.

12. The method of claim 1 wherein searching for a segment of the chunk that matches the memory block fails to find a matching segment, the method further comprising repeatedly carrying out the following steps for all memory blocks in computer memory until a matching segment is found:
- 5
- retrieving a next memory block from computer memory; and
- searching for a segment of the chunk that matches the next memory block.
13. The method of claim 12 wherein no matching segment is found in any memory block in computer memory, the method further comprising:
- storing a search section of the chunk; and
- 5
- providing a retrieval key for the search section of the chunk.
14. The method of claim 13 wherein storing a search section of the chunk comprises storing the search section of the chunk as a new memory block having a memory block size equal to the size of the search section of the chunk.
- 5
15. The method of claim 13 wherein providing a retrieval key for a search section of a chunk comprises:
- calculating a weak checksum for the search section of the chunk; and
- 5
- calculating a strong checksum for the search section of the chunk.
16. The method of claim 13 wherein providing a chunk comprises:
- retrieving from input a quantity of data equal in size to the search section; and

5 concatenating the retrieved input quantity to the remaining portion of the
chunk that remains after storing the search section.

17. The method of claim 1 further comprising:

receiving a retrieval key;

5 identifying a memory block in dependence upon the retrieval key; and

retrieving the identified memory block.

18. A system of content addressable data storage and compression for computer memory comprising:

means for providing a chunk of data comprising a quantity of input data;

5

means for retrieving a memory block from computer memory;

means for searching for a segment of the chunk that matches the memory block;

10

means for discarding a matching segment and providing a retrieval key for the memory block as a retrieval key for the matching segment;

means for identifying an unmatched portion of the chunk that does not match the memory block; and

15

means for storing the unmatched portion and providing a retrieval key for the unmatched portion.

19. The system of claim 18 wherein means for providing a chunk comprises means for retrieving from input a quantity of input data for storage and compression having a chunk size larger than a maximum memory block size.

20. The system of claim 18 wherein means for retrieving a memory block from computer memory comprises means for retrieving from computer memory a memory block having a memory block size no greater than a maximum memory block size.

5

21. The system of claim 18 wherein means for searching for a segment of the chunk that matches the memory block comprises means for searching at a repeating memory interval through a search section of the chunk for a segment

of the chunk that matches the memory block.

5

22. The system of claim 21 wherein the memory interval is one bit.

23. The system of claim 21 wherein means for searching at a repeating memory interval through a search section of the chunk for a segment of the chunk that matches the memory block comprises:

5

means for calculating a weak checksum for the memory block;

means for calculating weak checksums for segments of the search section of the chunk;

10

means for comparing the weak checksums for the segments with the checksum for the memory block;

means for calculating a strong checksum for the memory block;

15

means for calculating a strong checksum for the segment with the matching weak checksum;

means for comparing the strong checksum of the memory block and the strong checksum for the segment with the equal weak checksum;

20

means for determining that the search has found a segment having contents that match the contents of the memory block if the strong checksum of the memory block and the strong checksum for the segment with the matching weak checksum are equal.

25

24. The system of claim 23 wherein means for calculating a strong checksum for the memory block comprises means for calculating a static strong checksum

for the memory block.

25. The system of claim 23 wherein means for calculating weak checksums for segments of the search section of the chunk comprises means for calculating rolling weak checksums.

26. The system of claim 18 wherein means for storing the unmatched portion of the chunk comprises means for storing the unmatched portion of the chunk as a new memory block having a memory block size equal to the size of the unmatched portion of the chunk.

5

27. The system of claim 18 wherein means for providing a retrieval key for the unmatched portion of a chunk comprises:

5

means for calculating a weak checksum for the unmatched portion of the chunk; and

means for calculating a strong checksum for the unmatched portion of the chunk.

28. The system of claim 18 wherein providing a chunk comprises:

means for retrieving from input a quantity of data equal in size to the sum of the sizes of the matching segment and the unmatched portion; and

5

means for concatenating the retrieved input quantity to a remaining portion of the chunk that remains after discarding the matching segment and storing the unmatched portion.

29. The system of claim 18 wherein means for searching for a segment of the chunk that matches the memory block fails to find a matching segment, the

method further comprising:

- 5 means for retrieving a next memory block from computer memory; and
- means for searching for a segment of the chunk that matches the next memory block.

30. The system of claim 29 further comprising:

- means for storing a search section of the chunk; and
- 5 means for providing a retrieval key for the search section of the chunk.

31. The system of claim 30 wherein means for storing a search section of the chunk comprises means for storing the search section of the chunk as a new memory block having a memory block size equal to the size of the search section of the chunk.

- 5
32. The system of claim 30 wherein means for providing a retrieval key for a search section of a chunk comprises:

means for calculating a weak checksum for the search section of the chunk;

5 and

means for calculating a strong checksum for the search section of the chunk.

33. The system of claim 30 wherein means for providing a chunk comprises:

- means for retrieving from input a quantity of data equal in size to the search section; and
- 5

means for concatenating the retrieved input quantity to the remaining portion of the chunk that remains after storing the search section.

34. The system of claim 18 further comprising:

means for receiving a retrieval key;

5 means for identifying a memory block in dependence upon the retrieval key;
and

means for retrieving the identified memory block.

35. A computer program product of content addressable data storage and compression for computer memory comprising:

a recording medium;

5

means, recorded on the recording medium, for providing a chunk of data comprising a quantity of input data;

10

means, recorded on the recording medium, for retrieving a memory block from computer memory;

means, recorded on the recording medium, for searching for a segment of the chunk that matches the memory block;

15

means, recorded on the recording medium, for discarding a matching segment and providing a retrieval key for the memory block as a retrieval key for the matching segment;

20

means, recorded on the recording medium, for identifying an unmatched portion of the chunk that does not match the memory block; and

means, recorded on the recording medium, for storing the unmatched portion and providing a retrieval key for the unmatched portion.

36. The computer program product of claim 35 wherein means, recorded on the recording medium, for providing a chunk comprises means, recorded on the recording medium, for retrieving from input a quantity of input data for storage and compression having a chunk size larger than a maximum memory block size.

5

37. The computer program product of claim 35 wherein means, recorded on the

5 recording medium, for retrieving a memory block from computer memory comprises means, recorded on the recording medium, for retrieving from computer memory a memory block having a memory block size no greater than a maximum memory block size.

38. The computer program product of claim 35 wherein means, recorded on the recording medium, for searching for a segment of the chunk that matches the memory block comprises means, recorded on the recording medium, for searching at a repeating memory interval through a search section of the chunk for a segment of the chunk that matches the memory block.

39. The computer program product of claim 38 wherein the memory interval is one bit.

40. The computer program product of claim 38 wherein means, recorded on the recording medium, for searching at a repeating memory interval through a search section of the chunk for a segment of the chunk that matches the memory block comprises:

5 means, recorded on the recording medium, for calculating a weak checksum for the memory block;

10 means, recorded on the recording medium, for calculating weak checksums for segments of the search section of the chunk;

means, recorded on the recording medium, for comparing the weak checksums for the segments with the checksum for the memory block;

15 means, recorded on the recording medium, for calculating a strong checksum for the memory block;

means, recorded on the recording medium, for calculating a strong checksum for the segment with the matching weak checksum;

20

means, recorded on the recording medium, for comparing the strong checksum of the memory block and the strong checksum for the segment with the equal weak checksum;

25

means, recorded on the recording medium, for determining that the search has found a segment having contents that match the contents of the memory block if the strong checksum of the memory block and the strong checksum for the segment with the matching weak checksum are equal.

41. The computer program product of claim 40 wherein means, recorded on the recording medium, for calculating a strong checksum for the memory block comprises means, recorded on the recording medium, for calculating a static strong checksum for the memory block.

5

42. The computer program product of claim 40 wherein means, recorded on the recording medium, for calculating weak checksums for segments of the search section of the chunk comprises means, recorded on the recording medium, for calculating rolling weak checksums.

5

43. The computer program product of claim 35 wherein means, recorded on the recording medium, for storing the unmatched portion of the chunk comprises means, recorded on the recording medium, for storing the unmatched portion of the chunk as a new memory block having a memory block size equal to the size of the unmatched portion of the chunk.

5

44. The computer program product of claim 35 wherein means, recorded on the recording medium, for providing a retrieval key for the unmatched portion of a chunk comprises:

5 means, recorded on the recording medium, for calculating a weak checksum for the unmatched portion of the chunk; and

means, recorded on the recording medium, for calculating a strong checksum for the unmatched portion of the chunk.

10

45. The computer program product of claim 35 wherein providing a chunk comprises:

5 means, recorded on the recording medium, for retrieving from input a quantity of data equal in size to the sum of the sizes of the matching segment and the unmatched portion; and

10 means, recorded on the recording medium, for concatenating the retrieved input quantity to a remaining portion of the chunk that remains after discarding the matching segment and storing the unmatched portion.

46. The computer program product of claim 35 wherein means, recorded on the recording medium, for searching for a segment of the chunk that matches the memory block fails to find a matching segment, the method further comprising:

5

means, recorded on the recording medium, for retrieving a next memory block from computer memory; and

10 means, recorded on the recording medium, for searching for a segment of the chunk that matches the next memory block.

47. The computer program product of claim 46 further comprising:

means, recorded on the recording medium, for storing a search section of the chunk; and

5

means, recorded on the recording medium, for providing a retrieval key for the search section of the chunk.

48. The computer program product of claim 47 wherein means, recorded on the recording medium, for storing a search section of the chunk comprises means, recorded on the recording medium, for storing the search section of the chunk as a new memory block having a memory block size equal to the size of the search section of the chunk.

5

49. The computer program product of claim 47 wherein means, recorded on the recording medium, for providing a retrieval key for a search section of a chunk comprises:

5

means, recorded on the recording medium, for calculating a weak checksum for the search section of the chunk; and

means, recorded on the recording medium, for calculating a strong checksum for the search section of the chunk.

10

50. The computer program product of claim 47 wherein means, recorded on the recording medium, for providing a chunk comprises:

5

means, recorded on the recording medium, for retrieving from input a quantity of data equal in size to the search section; and

means, recorded on the recording medium, for concatenating the retrieved input quantity to the remaining portion of the chunk that remains after storing the search section.

10

51. The computer program product of claim 35 further comprising:

means, recorded on the recording medium, for receiving a retrieval key;

5 means, recorded on the recording medium, for identifying a memory block in dependence upon the retrieval key; and

means, recorded on the recording medium, for retrieving the identified memory block.